

PREDIKSI PENJUALAN PRODUK UNILEVER MENGGUNAKAN METODE K-NEAREST NEIGHBOR

Aisha Alfani W. P. R.¹⁾, Fahrur Rozi²⁾, Farid Sukmana³⁾

¹⁾ Jurusan Teknik Informatika, STMIK AKAKOM Yogyakarta
Jl. Raya Janti No.143, Jaranan, Yogyakarta

²⁾ Jurusan Teknologi Informasi, Universitas Bhinneka PGRI
Jl. Mayor Sujadi Timur No.7 Tulungagung

³⁾ Jurusan Teknik Informatika, Universitas Muhammadiyah Gresik
Jl. Sumatera No.101, Gn. Malang, Gresik

e-mail: aishaalfani03@gmail.com¹⁾, fahrur@ubhi.ac.id²⁾, faridsukmana@umg.ac.id³⁾

ABSTRAK

Toko Rizky Barokah Nganjuk merupakan sebuah toko retail semi grosir yang terletak di Kelurahan Bogo Kabupaten Nganjuk Jawa Timur. Toko ini menjual berbagai macam produk kebutuhan rumah tangga diantaranya dari Unilever, P&G, Nestlé, dan Indofood. Dilihat dari banyaknya permintaan konsumen akan produk Unilever berdasarkan data 3 tahun terakhir, maka dibutuhkan prediksi untuk penjualan produk Unilever terlaris. Hal ini berguna untuk mempermudah pihak pemilik toko dalam perencanaan penyediaan stok. Untuk mengetahui penjualan produk Unilever terlaris digunakan teknik klasifikasi data mining dan algoritma K-Nearest Neighbor. Hasil dari penelitian ini berdasarkan nilai akurasi tertinggi terhadap klasifikasi penjualan produk sebesar 86,66% dan nilai akurasi terendah terhadap klasifikasi penjualan produk sebesar 40%.

Kata kunci: forecasting, data mining, k-nearest neighbor (K-NN)

ABSTRACT

Rizky Barokah Nganjuk Store is a semi-wholesale retail store located in the Bogo District, Nganjuk, East Java. This store sells a variety of household products including Unilever, P&G, Nestlé and Indofood. Seen from the number of consumer requests for Unilever products based on data from the last 3 years, predictions are needed for sales of the best-selling Unilever products. This is useful to facilitate the store owner in planning the provision of stock. To find out the sales of the best-selling Unilever products, data mining classification techniques and the K-Nearest Neighbor algorithm are used. The results of this study are based on the highest accuracy value of the product sales classification of 86,66% and the lowest accuracy value of the product sales classification of 40%.

Keywords: forecasting, data mining, k-nearest neighbor (K-NN)

I. PENDAHULUAN

Prediksi atau peramalan penjualan (forecasting) merupakan suatu perhitungan untuk meramalkan keadaan di masa mendatang dengan melakukan pengujian terhadap keadaan di masa lalu [1]. Salah satu dari fungsi prediksi adalah untuk membantu pemilik perusahaan dalam pengambilan keputusan dan menentukan jumlah barang yang harus disediakan oleh perusahaan tersebut. Selain itu prediksi dapat membantu pihak perusahaan dalam penyediaan stok barang, karena prediksi ini dapat memberikan *output* terbaik sehingga diharapkan resiko kesalahan yang disebabkan oleh kesalahan perencanaan dapat ditekan seminimal mungkin. Prediksi biasanya digunakan untuk menemukan informasi dari data yang besar sehingga diperlukan *data mining*.

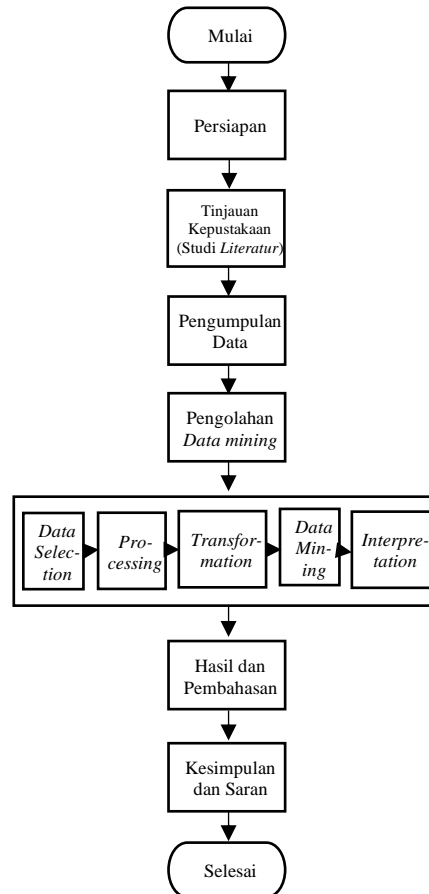
Data mining merupakan teknik pembelajaran mesin, pengenalan pola, database, statistic, dan visualisasi yang digunakan untuk penanganan masalah pengambilan informasi dari penyimpanan database yang besar [2]. Metode-metode yang terdapat dalam data mining untuk prediksi diantaranya yaitu metode K-Nearest Neighbor (K-NN). Metode ini digunakan untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut [3]. Penggunaan metode K-NN sebagai prediksi data sudah banyak digunakan, diantaranya K-NN dalam akitifitas manusia dalam pembelajaran online [4]. K-NN sebagai metode untuk klasifikasi terhadap data kanker [5].

Toko Rizky Barokah merupakan salah satu toko retail semi gorsir yang menyediakan barang-barang fast moving consumer goods (FMCG). Supplier toko ini antara lain Indofood, Nestlé, P&G, dan Unilever. Dengan begitu banyaknya barang yang dijual pada toko ini terdapat sebuah permasalahan yakni dalam sistem restock barang. Dilihat dari banyaknya permintaan konsumen akan produk-produk dari Unilever, maka dibutuhkan prediksi untuk

penjualan produk Unilever menggunakan metode K-NN.

II. METODE

Tahapan dalam penelitian ini dapat dilihat pada gambar 1 berikut [6]:.



Gambar 1. Tahapan Penelitian

Data yang digunakan dalam penelitian ini berasal dari Toko Rizky Barokah Nganjuk, yaitu data penjualan produk Unilever, P&G, Nestlé, dan Indofood dari tahun 2017, 2018, dan 2019 dengan range penjualan seperti pada tabel 1 berikut :

TABEL 1
RANGE PENJUALAN PRODUK

kategori	banyaknya penjualan (per minggu)
laris	> 200
sedang	80 - 200
kurang	< 80

Penelitian ini menggunakan metode pencarian jarak *euclidean distance* / perhitungan jarak terdekat. Perhitungan jarak terdekat dibutuhkan untuk menentukan jumlah kemiripan yang dihitung dari kemiripan kemunculan teks yang dimiliki suatu paragraph [7][8][9]. Setelah itu kemunculan teks yang sedang diujikan dibandingkan terhadap masing-masing sampel data asli [10][11]. Menghitung euclidean distance seperti pada persamaan (1) .

$$d_{(i,j)} = \sqrt{|x_{i1} - x_{j1}|^2 + \dots + |x_{ip} - x_{jp}|^2} \quad (1)$$

dimana, $d_{(i,j)}$ = Jarak data ke-i ke data ke-j, $x_{i(n)}$ = data latih , dan $x_{j(n)}$ = data uji. Confusion matrix adalah tool yang digunakan untuk evaluasi model klasifikasi untuk memperkirakan objek yang benar atau salah. Sebuah matrix

dari prediksi yang akan dibandingkan dengan kelas yang asli dari inputan atau dengan kata lain berisi informasi nilai aktual dan prediksi pada klasifikasi [12]. Adapun perhitungan tingkat akurasi pada confusion matrix ditunjukkan pada persamaan (2)[13][14] :

$$Akurasi = \frac{\Sigma \text{angka benar}}{\Sigma \text{angka pengujian}} \times 100\% \quad (2)$$

Dalam algoritma K-NN cara mengklasifikasikan output adalah berdasarkan besarnya nilai k. Pemilihan nilai k (tetangga terdekat) tidak memiliki standar yang baku, maka dalam penelitian ini nilai k yang digunakan adalah 10, 20, 30, dan 40. Berikut ini merupakan pengujian dengan hasil akurasi nilai k terendah dan nilai k tertinggi.

III. PEMBAHASAN

Contoh perhitungan manual dengan menggunakan K-Nearest Neighbor (KNN). Pada kasus ini menggunakan data penjualan produk yang dijual di Toko Rizky Barokah Nganjuk. Untuk memprediksi penjualan Unilever minggu ke-4 bulan Februari 2018, parameter yang digunakan adalah produk dari Indofood, P&G, dan Nestlé. Berikut ini adalah tabel 2 yang menunjukkan data penjualan produk Unilever di Toko Rizky Barokah Nganjuk.

TABEL II
PENJUALAN PRODUK UNILEVER DI TOKO RIZKY BAROKAH

Minggu ke-	Penjualan Produk Unilever
1	201
2	98
3	177
4	?

Berikut ini merupakan tabel penjualan barang yang akan dijadikan parameter untuk menghitung prediksi penjualan produk Unilever, seperti pada tabel 3:

TABEL III
PENJUALAN PRODUK DI TOKO RIZKY BAROKAH

Minggu Ke-	Indofood	Nestlé	P&G
1	125	102	131
2	143	115	111
3	93	91	178
4	112	87	193

Pada kasus ini nilai K yang digunakan adalah 2. Selanjutnya adalah Menghitung jarak euclidean data baru terhadap data yang ada menggunakan persamaan 1. Selanjutnya, Mengurutkan jarak euclidean terkecil. Table 4 menunjukkan urutan jarak euclidean terkecil.

TABEL IV
URUTAN EUCLIDEAN DISTANCE TERKECIL

Data Penjualan Februari 2018	<i>Squared distance to query distance</i> (170,112,87,193)	Urutan Jarak Terkecil	Kategori
Minggu ke-1 125 102 131 201	72,104	2	Sedang
Minggu ke-2 143 115 111 98	116,846	3	Sedang
Minggu ke-3 93 91 178 177	25,514	1	Kurang

Dari perhitungan diatas ditemukan hasil kategori sedang sebagai hasil yang dominan. Maka pada baris baru yang kategorinya masih dicari dapat dituliskan hasil sedang seperti pada tabel 5 berikut ini :

TABEL V
HASIL PERHITUNGAN

Minggu Ke-	Unilever	Indofood	Nestlé	P&G	Output
1	201	125	102	131	Sedang
2	98	143	115	111	Sedang
3	177	93	91	178	Kurang
4	?	112	87	193	Sedang

Dari table 4 diperoleh hasil “sedang” sebagai kategori yang mendominasi. Karena tetangga terdekat ($k=2$) hasil yang di dapat menunjukkan di klasifikasi sedang, maka dapat disimpulkan bahwa penjualan Produk Unilever di Toko Rizky Barokah pada Minggu ke-4 bulan Februari 2018 adalah dalam kategori sedang.

Pengujian yang dilakukan terhadap 50 data latih dan 10 data uji dengan nilai $k=10$ untuk mengetahui tingkat akurasi pada baris ke-51 penjualan produk Unilever ditunjukkan pada tabel 6.

TABEL VI
PENGUJIAN BARIS KE-51 ($K=10$)

penjualan ke-	output aktual	konversi output	output jaringan	konversi output
51	3	kurang	85	sedang
17	3	kurang	88	sedang
12	3	kurang	87	sedang
124	3	kurang	90	sedang
118	2	sedang	109	sedang
138	2	sedang	103	sedang
77	2	sedang	112	sedang
24	3	kurang	89	sedang
135	2	sedang	110	sedang
54	3	kurang	98	sedang

Tabel 1 merupakan hasil perhitungan dari 50 data latih (penjualan 1-50), dan 10 data uji (penjualan 51-60) dengan nilai $k = 10$. Dapat dilihat bahwa dari 10 data diatas antara output aktual dengan output jaringan menghasilkan 6 output yang berbeda. Maka dapat disimpulkan bahwa pengujian terhadap bilangan tersebut memiliki tingkat kearurasian sebesar 40%.

Pengujian yang dilakukan terhadap 60 data latih dan 20 data uji dengan nilai $k=30$ untuk mengetahui tingkat akurasi pada baris ke-80 penjualan produk Unilever ditunjukkan pada tabel 7.

TABEL VII
PENGUJIAN BARIS KE-80 ($K=30$)

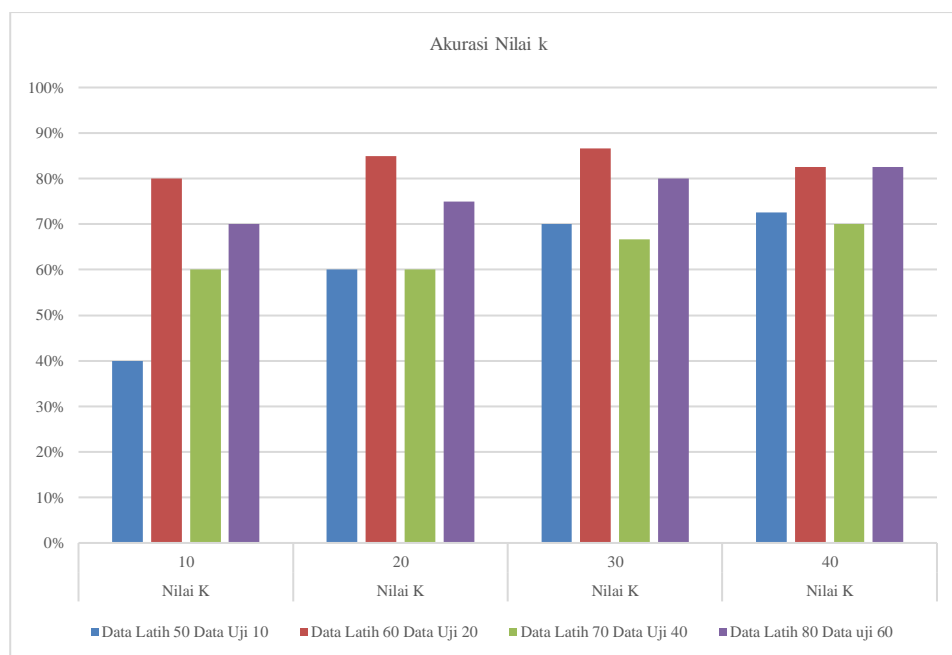
penjualan ke-	output aktual	konversi output	output jaringan	konversi output
80	2	sedang	173	sedang
115	2	sedang	169	sedang
117	2	sedang	189	sedang
76	2	sedang	192	sedang
121	2	sedang	173	sedang
104	2	sedang	209	laris
120	2	sedang	178	sedang
65	2	sedang	212	laris
107	2	sedang	170	sedang
25	2	sedang	164	sedang
116	2	sedang	130	sedang
127	2	sedang	168	sedang
20	2	sedang	172	sedang
28	2	sedang	172	sedang

penjualan ke-	output aktual	konversi output	output jaringan	konversi output
13	2	sedang	165	sedang
41	2	sedang	190	sedang
22	2	sedang	171	sedang
39	2	sedang	218	laris
72	2	sedang	130	sedang
106	2	sedang	155	sedang
122	2	sedang	193	sedang
86	2	sedang	190	sedang
27	2	sedang	150	sedang
109	2	sedang	143	sedang
37	2	sedang	222	laris
44	2	sedang	139	sedang
56	2	sedang	170	sedang
71	2	sedang	152	sedang
6	2	sedang	189	sedang
4	2	sedang	111	sedang

Tabel 2 merupakan hasil perhitungan dari 60 data latih (penjualan 1-60), dan 20 data uji (penjualan 61-80) dengan nilai $k = 30$. Dapat dilihat bahwa dari 30 data diatas antara output aktual dengan output jaringan menghasilkan 4 output yang berbeda. Maka dapat disimpulkan bahwa pengujian terhadap bilangan tersebut memiliki tingkat kearurasian 86,66%.

Berdasarkan pengujian yang telah dilakukan pada seluruh nilai k , didapatkan hasil adalah akurasi tertinggi nilai k sebesar 86,66% yang dihasilkan pada pengujian baris ke-80 dengan nilai $k=30$. Akurasi terendah nilai k sebesar 40% yang dihasilkan pada pengujian baris ke-51 dengan nilai $k=10$. Nilai k yang menghasilkan output konsisten meliputi 10, 20, dan 30. Hal ini dibuktikan bahwa seluruh angka yang diuji menggunakan nilai k tersebut menghasilkan output yang semakin besar tingkat akurasinya. Nilai k yang menghasilkan penurunan tingkat akurasi pada output yang dihasilkan adalah $k=40$. Hal ini dibuktikan pada pengujian baris ke-80 dan 135 tingkat akurasi nilai k mengalami penurunan. Penyebab dari penurunan tingkat akurasi nilai $k=40$ adalah apabila nilai k yang digunakan pada suatu pengujian > 30 , maka nilai yang dihasilkan akan semakin heterogen.

Diagram berikut ini merupakan gambaran dari hasil akhir nilai k yang telah diuji. Terdapat 4 buah nilai k yang digunakan peneliti dalam peneltian ini yaitu 10, 20, 30, 40. Berikut merupakan gambar 2 yang menampilkan perbandingan akurasi nilai k .



Gambar 2. Diagram Batang Akurasi Nilai k

Sehingga dapat ditarik kesimpulan bahwa pemilihan nilai k yang tepat untuk digunakan dalam sebuah algoritma K-NN mempengaruhi nilai akurasi secara signifikan. Hal ini dibuktikan bahwa hasil perhitungan euclidean distance antara output aktual dan output jaringan yang diperoleh dalam pengujian tersebut adalah fluktuatif.

IV. KESIMPULAN

Berdasarkan hasil perhitungan data mining menggunakan teknik klasifikasi dan algoritma K-NN, didapatkan hasil prediksi penjualan produk berdasarkan nilai akurasi tertinggi dan terendah. Nilai akurasi tertinggi terhadap klasifikasi penjualan produk sebesar 86,66%. Sedangkan nilai akurasi terendah terhadap klasifikasi penjualan produk sebesar 40%. Dengan demikian metode data mining dan algoritma K-NN ini dapat diimplementasikan untuk memprediksi penjualan produk Unilever di Toko Rizky Barokah Nganjuk.

DAFTAR PUSTAKA

- [1] F. Rozi and R. Kartadie, "Deteksi E-Mail Dan Spam Menggunakan Fuzzy Association Rule Mining," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 2, no. 2, pp. 94–98, 2017, doi: 10.29100/jipi.v2i2.348.
- [2] T. B. Alakus and I. Turkoglu, "Comparison of deep learning approaches to predict COVID-19 infection," *Chaos, Solitons and Fractals*, vol. 140, p. 110120, 2020, doi: 10.1016/j.chaos.2020.110120.
- [3] A. Ali, M. Alrubei, L. F. M. Hassan, M. Al-Ja'afari, and S. Abdulwahed, "Diabetes classification based on KNN," *IJUM Eng. J.*, 2020, doi: 10.31436/iiumej.v2i1i1.1206.
- [4] P. J. S. Ferreira, J. M. P. Cardoso, and J. Mendes-Moreira, "KNN prototyping schemes for embedded human activity recognition with online learning," *Computers*, 2020, doi: 10.3390/computers9040096.
- [5] R. Ehsani and F. Drabløs, "Robust Distance Measures for kNN Classification of Cancer Data," *Cancer Inform.*, 2020, doi: 10.1177/1176935120965542.
- [6] J. Hu, H. Peng, J. Wang, and W. Yu, "kNN-P: A kNN classifier optimized by P systems," *Theor. Comput. Sci.*, 2020, doi: 10.1016/j.tcs.2020.01.001.
- [7] Z. Chen, L. J. Zhou, X. Da Li, J. N. Zhang, and W. J. Huo, "The Lao text classification method based on KNN," in *Procedia Computer Science*, 2020, doi: 10.1016/j.procs.2020.02.053.
- [8] F. Rozi and R. Kartadie, "Sinonim untuk ekstraksi kata kunci pada pengelompokan dokumen menggunakan fuzzy association rule mining," *Semnasteknomedia Online*, pp. 7–12, 2016.
- [9] F. Rozi, C. Faticah, and D. Purwitasari, "Ekstraksi Kata Kunci Berdasarkan Hipernim Menggunakan Fuzzy Association Rule Mining untuk Pengelompokan Dokumen," *J. Ilm. Teknol. Inf.*, vol. 13, no. 2, pp. 190–197, 2015.
- [10] J. Hartmann, J. Huppertz, C. Schamp, and M. Heitmann, "Comparing automated text classification methods," *Int. J. Res. Mark.*, 2019, doi: 10.1016/j.ijresmar.2018.09.009.
- [11] F. Sukmana and F. Rozi, "Decision support system on computer maintenance management system using Association Rule and Fisher Exact test One Side P-Value," *Telkomnika (Telecommunication Comput. Electron. Control.*, vol. 15, no. 4, 2017, doi: 10.12928/TELKOMNIKA.v15i4.5880.
- [12] D. Ö. Şahin and E. Kılıç, "Two new feature selection metrics for text classification," *Automatika*, 2019, doi: 10.1080/00051144.2019.1602293.
- [13] S. Xu, "Bayesian Naïve Bayes classifiers to text classification," *J. Inf. Sci.*, 2018, doi: 10.1177/0165551516677946.
- [14] F. Sukmana and F. Rozi, "Extraction keyterm in work order for decision support," *J. Theor. Appl. Inf. Technol.*, vol. 97, no. 22, pp. 3262–3272, 2019.